



Munich Personal RePEc Archive

Loss Distributions

Krzysztof Burnecki and Adam Misiorek and Rafal Weron

2010

Online at <http://mpa.ub.uni-muenchen.de/22163/>

MPRA Paper No. 22163, posted 20. April 2010 10:14 UTC

Loss Distributions¹

Krzysztof Burnecki^a, Adam Misiorek^{b,c}, Rafał Weron^b

^a*Hugo Steinhaus Center, Wrocław University of Technology, 50-370 Wrocław, Poland*

^b*Institute of Organization and Management, Wrocław University of Technology, 50-370 Wrocław, Poland*

^c*Santander Consumer Bank S.A., Wrocław, Poland*

Abstract

This paper is intended as a guide to statistical inference for loss distributions. There are three basic approaches to deriving the loss distribution in an insurance risk model: empirical, analytical, and moment based. The empirical method is based on a sufficiently smooth and accurate estimate of the cumulative distribution function (cdf) and can be used only when large data sets are available. The analytical approach is probably the most often used in practice and certainly the most frequently adopted in the actuarial literature. It reduces to finding a suitable analytical expression which fits the observed data well and which is easy to handle. In some applications the exact shape of the loss distribution is not required. We may then use the moment based approach, which consists of estimating only the lowest characteristics (moments) of the distribution, like the mean and variance.

Having a large collection of distributions to choose from, we need to narrow our selection to a single model and a unique parameter estimate. The type of the objective loss distribution can be easily selected by comparing the shapes of the empirical and theoretical mean excess functions. Goodness-of-fit can be verified by plotting the corresponding limited expected value functions. Finally, the hypothesis that the modeled random event is governed by a certain loss distribution can be statistically tested.

Keywords: Loss distribution, Insurance risk model, Random variable generation, Goodness-of-fit testing, Mean excess function, Limited expected value function.

1. Introduction

The derivation of loss distributions from insurance data is not an easy task. Insurers normally keep data files containing detailed information about policies and claims, which are used for accounting and rate-making purposes. However, claim size distributions and other data needed for risk-theoretical analyzes can be obtained usually only after tedious data preprocessing. Moreover, the claim statistics are often limited. Data files containing detailed information about some policies and claims may be missing or corrupted. There may also be situations where prior data or experience are not available at all, e.g. when a new type of insurance is introduced or when very large special risks are insured. Then the distribution has to be based on knowledge of similar risks or on extrapolation of lesser risks.

There are three basic approaches to deriving the loss distribution: empirical, analytical, and moment based. The empirical method, presented in Section 2, can be used only when large data sets are available. In such cases a sufficiently smooth and accurate estimate of the cumulative distribution function (cdf) is obtained. Sometimes the application of curve fitting techniques – used to smooth the empirical distribution function – can be beneficial. If the curve can be described by a function with a tractable analytical form, then this approach becomes computationally efficient and similar to the second method.

The analytical approach is probably the most often used in practice and certainly the most frequently adopted in the actuarial literature. It reduces to finding a suitable analytical expression which fits the observed data well and which is easy to handle. Basic characteristics and estimation issues for the most popular and useful loss distributions

¹This is a revised version of a paper originally published as Chapter 13 *Loss Distributions* in Čižek, Härdle, and Weron (2005). Among other changes, the revision includes new figures (thanks go to Joanna Janczura) and corrected formulas.

are discussed in Section 3. Note, that sometimes it may be helpful to subdivide the range of the claim size distribution into intervals for which different methods are employed. For example, the small and medium size claims could be described by the empirical claim size distribution, while the large claims – for which the scarcity of data eliminates the use of the empirical approach – by an analytical loss distribution.

In some applications the exact shape of the loss distribution is not required. We may then use the moment based approach, which consists of estimating only the lowest characteristics (moments) of the distribution, like the mean and variance. However, it should be kept in mind that even the lowest three or four moments do not fully define the shape of a distribution, and therefore the fit to the observed data may be poor. Further details on the moment based approach can be found e.g. in Daykin, Pentikainen, and Pesonen (1994).

Having a large collection of distributions to choose from, we need to narrow our selection to a single model and a unique parameter estimate. The type of the objective loss distribution can be easily selected by comparing the shapes of the empirical and theoretical mean excess functions. Goodness-of-fit can be verified by plotting the corresponding limited expected value functions. Finally, the hypothesis that the modeled random event is governed by a certain loss distribution can be statistically tested. In Section 4 these statistical issues are thoroughly discussed.

In Section 5 we apply the presented tools to modeling real-world insurance data. The analysis is conducted for two datasets: (i) the PCS (Property Claim Services) dataset covering losses resulting from catastrophic events in USA that occurred between 1990 and 1999 and (ii) the Danish fire losses dataset, which concerns major fire losses that occurred between 1980 and 1990 and were recorded by Copenhagen Re.

2. Empirical distribution function

A natural estimate for the loss distribution is the observed (empirical) claim size distribution. However, if there have been changes in monetary values during the observation period, inflation corrected data should be used. For a sample of observations $\{x_1, \dots, x_n\}$ the empirical distribution function (edf) is defined as:

$$F_n(x) = \frac{1}{n} \#\{i : x_i \leq x\}, \quad (1)$$

i.e. it is a piecewise constant function with jumps of size $1/n$ at points x_i . Very often, especially if the sample is large, the edf is approximated by a continuous, piecewise linear function with the “jump points” connected by linear functions, see Figure 1.

The empirical distribution function approach is appropriate only when there is a sufficiently large volume of claim data. This is rarely the case for the tail of the distribution, especially in situations where exceptionally large claims are possible. It is often advisable to divide the range of relevant values of claims into two parts, treating the claim sizes up to some limit on a discrete basis, while the tail is replaced by an analytical cdf.

3. Analytical methods

It is often desirable to find an explicit analytical expression for a loss distribution. This is particularly the case if the claim statistics are too sparse to use the empirical approach. It should be stressed, however, that many standard models in statistics – like the Gaussian distribution – are unsuitable for fitting the claim size distribution. The main reason for this is the strongly skewed nature of loss distributions. The log-normal, Pareto, Burr, Weibull, and gamma distributions are typical candidates for claim size distributions to be considered in applications.

3.1. Log-normal distribution

Consider a random variable X which has the normal distribution with density

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right\}, \quad -\infty < x < \infty. \quad (2)$$

Let $Y = e^X$ so that $X = \log Y$. Then the probability density function of Y is given by:

$$f(y) = f_N(\log y) \frac{1}{y} = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{1}{2} \frac{(\log y - \mu)^2}{\sigma^2}\right\}, \quad y > 0, \quad (3)$$

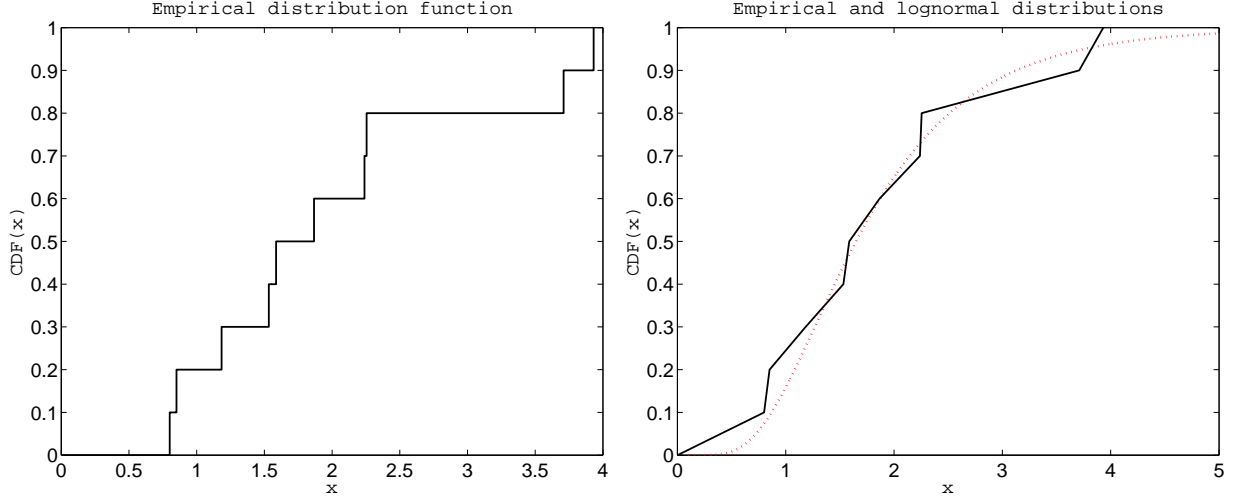


Figure 1: *Left panel:* Empirical distribution function (edf) of a 10-element log-normally distributed sample with parameters $\mu = 0.5$ and $\sigma = 0.5$, see Section 3.1. *Right panel:* Approximation of the edf by a continuous, piecewise linear function (black solid line) and the theoretical distribution function (red dotted line).

where $\sigma > 0$ is the scale and $-\infty < \mu < \infty$ is the location parameter. The distribution of Y is termed log-normal, however, sometimes it is called the Cobb-Douglas law, especially when applied to econometric data. The log-normal cdf is given by:

$$F(y) = \Phi\left(\frac{\log y - \mu}{\sigma}\right), \quad y > 0, \quad (4)$$

where $\Phi(\cdot)$ is the standard normal (with mean 0 and variance 1) distribution function. The k -th raw moment m_k of the log-normal variate can be easily derived using results for normal random variables:

$$m_k = E(Y^k) = E(e^{kX}) = M_X(k) = \exp\left(\mu k + \frac{\sigma^2 k^2}{2}\right), \quad (5)$$

where $M_X(z)$ is the moment generating function of the normal distribution. In particular, the mean and variance are

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad (6)$$

$$\text{Var}(X) = \{\exp(\sigma^2) - 1\} \exp(2\mu + \sigma^2), \quad (7)$$

respectively. For both standard parameter estimation techniques the estimators are known in closed form. The method of moments estimators are given by:

$$\hat{\mu} = 2 \log\left(\frac{1}{n} \sum_{i=1}^n x_i\right) - \frac{1}{2} \log\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right), \quad (8)$$

$$\hat{\sigma}^2 = \log\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - 2 \log\left(\frac{1}{n} \sum_{i=1}^n x_i\right), \quad (9)$$

while the maximum likelihood estimators by:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log(x_i), \quad (10)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \{\log(x_i) - \hat{\mu}\}^2. \quad (11)$$

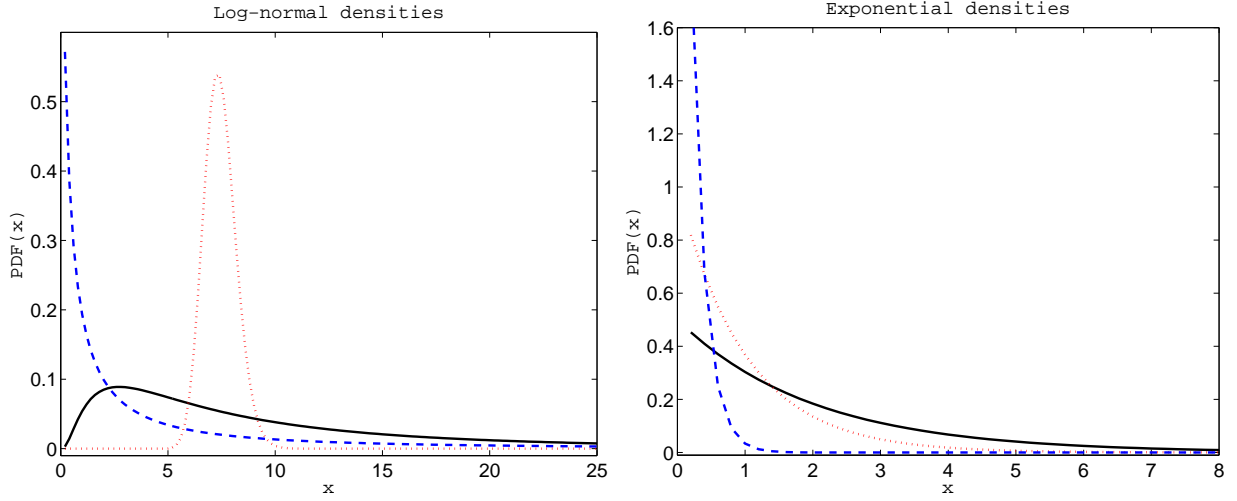


Figure 2: *Left panel:* Log-normal probability density functions (pdfs) with parameters $\mu = 2$ and $\sigma = 1$ (black solid line), $\mu = 2$ and $\sigma = 0.1$ (red dotted line), and $\mu = 0.5$ and $\sigma = 2$ (blue dashed line). *Right panel:* Exponential pdfs with parameter $\beta = 0.5$ (black solid line), $\beta = 1$ (red dotted line), and $\beta = 5$ (blue dashed line).

Finally, the generation of a log-normal variate is straightforward. We simply have to take the exponent of a normal variate.

The log-normal distribution is very useful in modeling of claim sizes. It is right-skewed, has a thick tail and fits many situations well. For small σ it resembles a normal distribution (see the left panel in Figure 2) although this is not always desirable. It is infinitely divisible and closed under scale and power transformations. However, it also suffers from some drawbacks. Most notably, the Laplace transform does not have a closed form representation and the moment generating function does not exist.

3.2. Exponential distribution

Consider the random variable with the following density and distribution functions, respectively:

$$f(x) = \beta e^{-\beta x}, \quad x > 0, \quad (12)$$

$$F(x) = 1 - e^{-\beta x}, \quad x > 0. \quad (13)$$

This distribution is termed an exponential distribution with parameter (or intensity) $\beta > 0$. The Laplace transform of (12) is

$$L(t) \stackrel{\text{def}}{=} \int_0^\infty e^{-tx} f(x) dx = \frac{\beta}{\beta + t}, \quad t > -\beta, \quad (14)$$

yielding the general formula for the k -th raw moment

$$m_k \stackrel{\text{def}}{=} (-1)^k \frac{\partial^k L(t)}{\partial t^k} \Big|_{t=0} = \frac{k!}{\beta^k}. \quad (15)$$

The mean and variance are thus β^{-1} and β^{-2} , respectively. The maximum likelihood estimator (equal to the method of moments estimator) for β is given by:

$$\hat{\beta} = \frac{1}{\hat{m}_1}, \quad (16)$$

where

$$\hat{m}_k = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad (17)$$

is the sample k -th raw moment.

To generate an exponential random variable X with intensity β we can use the inverse transform method (L'Ecuyer, 2004; Ross, 2002). The method consists of taking a random number U distributed uniformly on the interval $(0, 1)$ and setting $X = F^{-1}(U)$, where $F^{-1}(x) = -\frac{1}{\beta} \log(1 - x)$ is the inverse of the exponential cdf (13). In fact we can set $X = -\frac{1}{\beta} \log U$ since $1 - U$ has the same distribution as U .

The exponential distribution has many interesting features. For example, it has the memoryless property, i.e. $P(X > x + y | X > y) = P(X > x)$. It also arises as the inter-occurrence times of the events in a Poisson process, see Chapter 14 in Cížek, Härdle, and Weron (2005). The n -th root of the Laplace transform (14) is

$$L(t) = \left(\frac{\beta}{\beta + t} \right)^{\frac{1}{n}}, \quad (18)$$

which is the Laplace transform of a gamma variate (see Section 3.6). Thus the exponential distribution is infinitely divisible.

The exponential distribution is often used in developing models of insurance risks. This usefulness stems in a large part from its many and varied tractable mathematical properties. However, a disadvantage of the exponential distribution is that its density is monotone decreasing (see the right panel in Figure 2), a situation which may not be appropriate in some practical situations.

3.3. Pareto distribution

Suppose that a variate X has (conditional on β) an exponential distribution with mean β^{-1} . Further, suppose that β itself has a gamma distribution (see Section 3.6). The unconditional distribution of X is a mixture and is called the Pareto distribution. Moreover, it can be shown that if X is an exponential random variable and Y is a gamma random variable, then X/Y is a Pareto random variable.

The density and distribution functions of a Pareto variate are given by:

$$f(x) = \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > 0, \quad (19)$$

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x} \right)^\alpha, \quad x > 0, \quad (20)$$

respectively. Clearly, the shape parameter α and the scale parameter λ are both positive. The k -th raw moment:

$$m_k = \lambda^k k! \frac{\Gamma(\alpha - k)}{\Gamma(\alpha)}, \quad (21)$$

exists only for $k < \alpha$. In the above formula

$$\Gamma(a) \stackrel{\text{def}}{=} \int_0^\infty y^{a-1} e^{-y} dy, \quad (22)$$

is the standard gamma function. The mean and variance are thus:

$$E(X) = \frac{\lambda}{\alpha - 1}, \quad (23)$$

$$\text{Var}(X) = \frac{\alpha \lambda^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad (24)$$

respectively. Note, that the mean exists only for $\alpha > 1$ and the variance only for $\alpha > 2$. Hence, the Pareto distribution has very thick (or heavy) tails, see Figure 3. The method of moments estimators are given by:

$$\hat{\alpha} = 2 \frac{\hat{m}_2 - \hat{m}_1^2}{\hat{m}_2 - 2\hat{m}_1^2}, \quad (25)$$

$$\hat{\lambda} = \frac{\hat{m}_1 \hat{m}_2}{\hat{m}_2 - 2\hat{m}_1^2}, \quad (26)$$

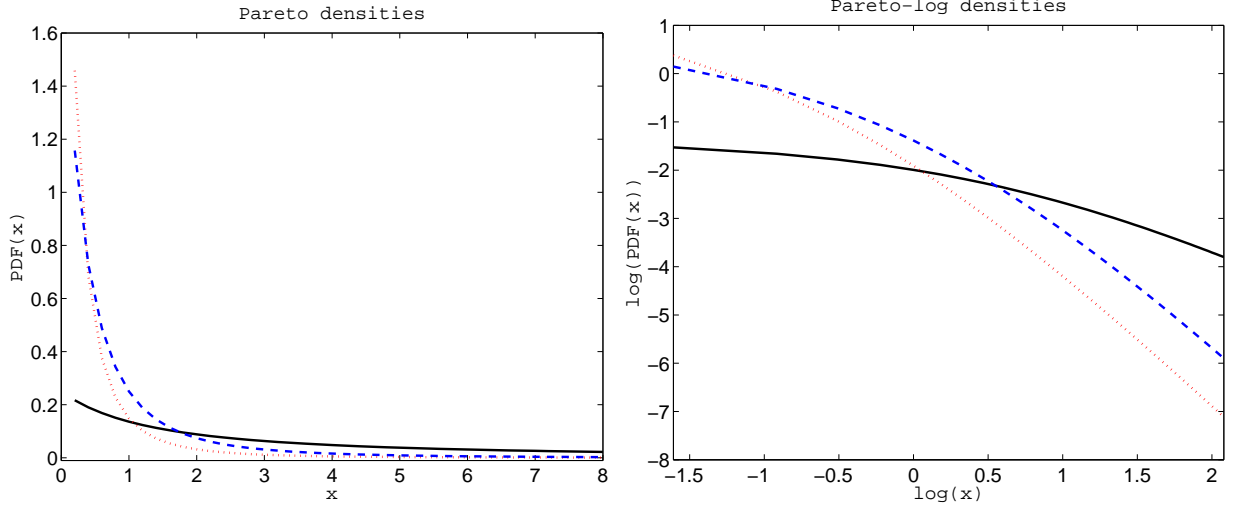


Figure 3: *Left panel:* Pareto pdfs with parameters $\alpha = 0.5$ and $\lambda = 2$ (black solid line), $\alpha = 2$ and $\lambda = 0.5$ (red dotted line), and $\alpha = 2$ and $\lambda = 1$ (blue dashed line). *Right panel:* The same Pareto densities on a double logarithmic plot. The thick power-law tails of the Pareto distribution are clearly visible.

where, as before, \hat{m}_k is the sample k -th raw moment (17). Note, that the estimators are well defined only when $\hat{m}_2 - 2\hat{m}_1^2 > 0$. Unfortunately, there are no closed form expressions for the maximum likelihood estimators and they can only be evaluated numerically.

Like for many other distributions the simulation of a Pareto variate X can be conducted via the inverse transform method. The inverse of the cdf (20) has a simple analytical form $F^{-1}(x) = \lambda \{(1-x)^{-1/\alpha} - 1\}$. Hence, we can set $X = \lambda(U^{-1/\alpha} - 1)$, where U is distributed uniformly on the unit interval. We have to be cautious, however, when α is larger but very close to one. The theoretical mean exists, but the right tail is very heavy. The sample mean will, in general, be significantly lower than $E(X)$.

The Pareto law is very useful in modeling claim sizes in insurance, due in large part to its extremely thick tail. Its main drawback lies in its lack of mathematical tractability in some situations. Like for the log-normal distribution, the Laplace transform does not have a closed form representation and the moment generating function does not exist. Moreover, like the exponential pdf the Pareto density (19) is monotone decreasing, which may not be adequate in some practical situations.

3.4. Burr distribution

Experience has shown that the Pareto formula is often an appropriate model for the claim size distribution, particularly where exceptionally large claims may occur. However, there is sometimes a need to find heavy tailed distributions which offer greater flexibility than the Pareto law, including a non-monotone pdf. Such flexibility is provided by the Burr distribution and its additional shape parameter $\tau > 0$. If Y has the Pareto distribution, then the distribution of $X = Y^{1/\tau}$ is known as the Burr distribution, see the left panel in Figure 4. Its density and distribution functions are given by:

$$f(x) = \tau\alpha\lambda^\alpha \frac{x^{\tau-1}}{(\lambda + x^\tau)^{\alpha+1}}, \quad x > 0, \quad (27)$$

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x^\tau}\right)^\alpha, \quad x > 0, \quad (28)$$

respectively. The k -th raw moment

$$m_k = \frac{1}{\Gamma(\alpha)} \lambda^{k/\tau} \Gamma\left(1 + \frac{k}{\tau}\right) \Gamma\left(\alpha - \frac{k}{\tau}\right), \quad (29)$$

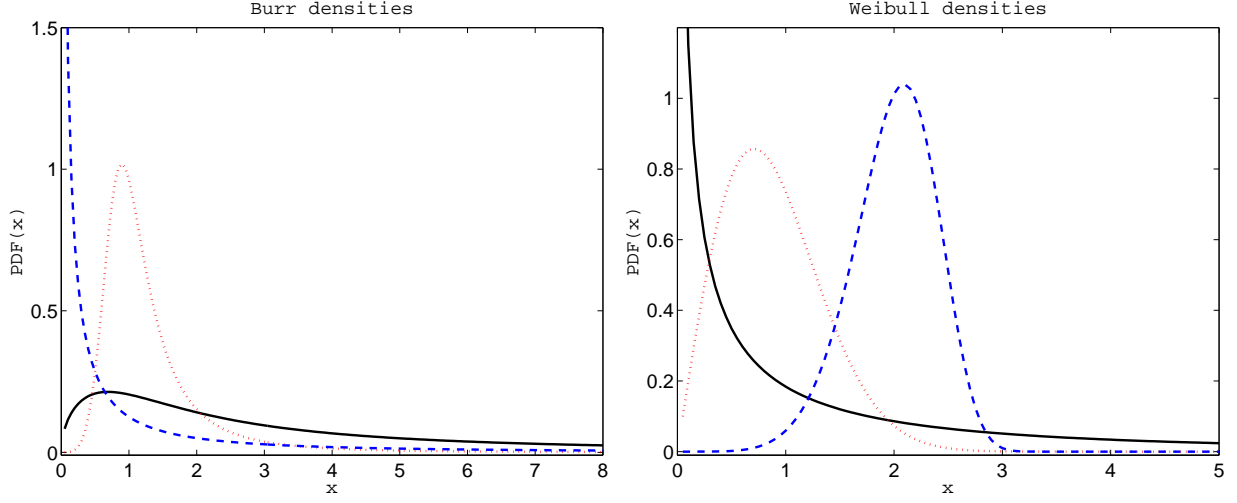


Figure 4: *Left panel:* Burr pdfs with parameters $\alpha = 0.5$, $\lambda = 2$ and $\tau = 1.5$ (black solid line), $\alpha = 0.5$, $\lambda = 0.5$ and $\tau = 5$ (red dotted line), and $\alpha = 2$, $\lambda = 1$ and $\tau = 0.5$ (blue dashed line). *Right panel:* Weibull pdfs with parameters $\beta = 1$ and $\tau = 0.5$ (black solid line), $\beta = 1$ and $\tau = 2$ (red dotted line), and $\beta = 0.01$ and $\tau = 6$ (blue dashed line).

exists only for $k < \tau\alpha$. Naturally, the Laplace transform does not exist in a closed form and the distribution has no moment generating function as it was the case with the Pareto distribution.

The maximum likelihood and method of moments estimators for the Burr distribution can only be evaluated numerically. A Burr variate X can be generated using the inverse transform method. The inverse of the cdf (28) has a simple analytical form $F^{-1}(x) = \left[\lambda \left\{ (1-x)^{-1/\alpha} - 1 \right\} \right]^{1/\tau}$. Hence, we can set $X = \left\{ \lambda \left(U^{-1/\alpha} - 1 \right) \right\}^{1/\tau}$, where U is distributed uniformly on the unit interval. Like in the Pareto case, we have to be cautious when $\tau\alpha$ is larger but very close to one. The theoretical mean exists, but the right tail is very heavy. The sample mean will, in general, be significantly lower than $E(X)$.

3.5. Weibull distribution

If V is an exponential variate, then the distribution of $X = V^{1/\tau}$, $\tau > 0$, is called the Weibull (or Fréchet) distribution. Its density and distribution functions are given by:

$$f(x) = \tau \beta x^{\tau-1} e^{-\beta x^\tau}, \quad x > 0, \quad (30)$$

$$F(x) = 1 - e^{-\beta x^\tau}, \quad x > 0, \quad (31)$$

respectively. The Weibull distribution is roughly symmetrical for the shape parameter $\tau \approx 3.6$. When τ is smaller the distribution is right-skewed, when τ is larger it is left-skewed, see the right panel in Figure 4. The k -th raw moment can be shown to be

$$m_k = \beta^{-k/\tau} \Gamma\left(1 + \frac{k}{\tau}\right). \quad (32)$$

Like for the Burr distribution, the maximum likelihood and method of moments estimators can only be evaluated numerically. Similarly, Weibull variates can be generated using the inverse transform method.

3.6. Gamma distribution

The probability law with density and distribution functions given by:

$$f(x) = \beta(\beta x)^{\alpha-1} \frac{e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \quad (33)$$

$$F(x) = \int_0^x \beta(\beta s)^{\alpha-1} \frac{e^{-\beta s}}{\Gamma(\alpha)} ds, \quad x > 0, \quad (34)$$

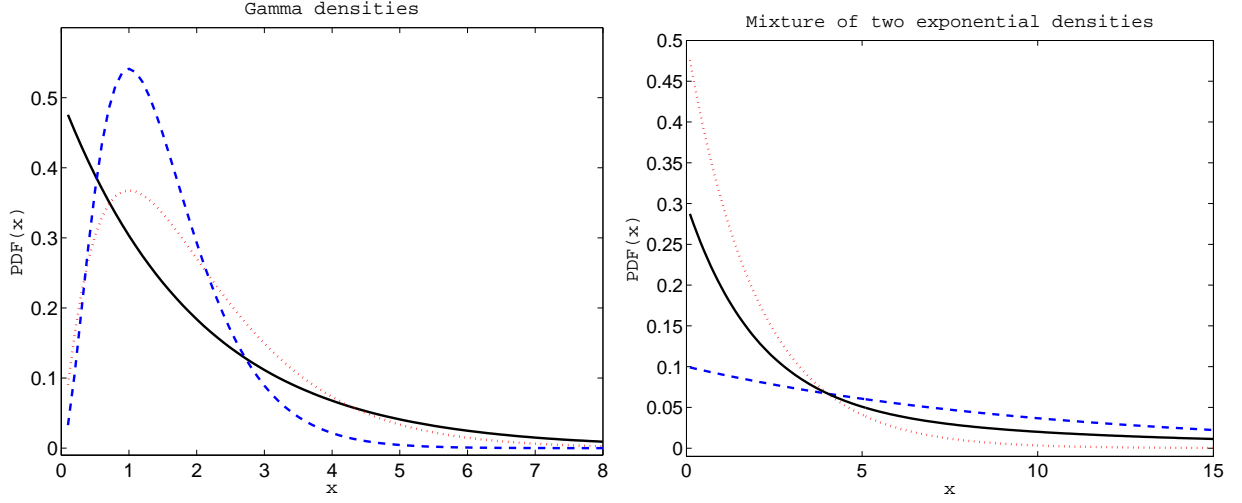


Figure 5: *Left panel:* Gamma pdfs with parameters $\alpha = 1$ and $\beta = 2$ (black solid line), $\alpha = 2$ and $\beta = 1$ (red dotted line), and $\alpha = 3$ and $\beta = 0.5$ (blue dashed line). *Right panel:* Densities of two exponential distributions with parameters $\beta_1 = 0.5$ (red dotted line) and $\beta_2 = 0.1$ (blue dashed line) and of their mixture with the mixing parameter $a = 0.5$ (black solid line).

where α and β are non-negative, is known as a gamma (or a Pearson's Type III) distribution, see the left panel in Figure 5. Moreover, for $\beta = 1$ the integral in (34):

$$\Gamma(\alpha, x) \stackrel{\text{def}}{=} \frac{1}{\Gamma(\alpha)} \int_0^x s^{\alpha-1} e^{-s} ds, \quad (35)$$

is called the incomplete gamma function. If the shape parameter $\alpha = 1$, the exponential distribution results. If α is a positive integer, the distribution is termed an Erlang law. If $\beta = \frac{1}{2}$ and $\alpha = \frac{\nu}{2}$ then it is termed a chi-squared (χ^2) distribution with ν degrees of freedom. Moreover, a mixed Poisson distribution with gamma mixing distribution is negative binomial, see Chapter 18 in Cížek, Härdle, and Weron (2005).

The Laplace transform of the gamma distribution is given by:

$$L(t) = \left(\frac{\beta}{\beta + t} \right)^\alpha, \quad t > -\beta. \quad (36)$$

The k -th raw moment can be easily derived from the Laplace transform:

$$m_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha) \beta^k}. \quad (37)$$

Hence, the mean and variance are

$$E(X) = \frac{\alpha}{\beta}, \quad (38)$$

$$\text{Var}(X) = \frac{\alpha}{\beta^2}. \quad (39)$$

Finally, the method of moments estimators for the gamma distribution parameters have closed form expressions:

$$\hat{\alpha} = \frac{\hat{m}_1^2}{\hat{m}_2 - \hat{m}_1^2}, \quad (40)$$

$$\hat{\beta} = \frac{\hat{m}_1}{\hat{m}_2 - \hat{m}_1^2}, \quad (41)$$

but maximum likelihood estimators can only be evaluated numerically. Simulation of gamma variates is not as straightforward as for the distributions presented above. For $\alpha < 1$ a simple but slow algorithm due to Jönlk (1964) can be used, while for $\alpha > 1$ the rejection method is more optimal (Bratley, Fox, and Schrage, 1987; Devroye, 1986).

The gamma distribution is closed under convolution, i.e. a sum of independent gamma variates with the same parameter β is again gamma distributed with this β . Hence, it is infinitely divisible. Moreover, it is right-skewed and approaches a normal distribution in the limit as α goes to infinity.

The gamma law is one of the most important distributions for modeling because it has very tractable mathematical properties. As we have seen above it is also very useful in creating other distributions, but by itself is rarely a reasonable model for insurance claim sizes.

3.7. Mixture of exponential distributions

Let a_1, a_2, \dots, a_n denote a series of non-negative weights satisfying $\sum_{i=1}^n a_i = 1$. Let $F_1(x), F_2(x), \dots, F_n(x)$ denote an arbitrary sequence of exponential distribution functions given by the parameters $\beta_1, \beta_2, \dots, \beta_n$, respectively. Then, the distribution function:

$$F(x) = \sum_{i=1}^n a_i F_i(x) = \sum_{i=1}^n a_i \{1 - \exp(-\beta_i x)\}, \quad (42)$$

is called a mixture of n exponential distributions (exponentials). The density function of the constructed distribution is

$$f(x) = \sum_{i=1}^n a_i f_i(x) = \sum_{i=1}^n a_i \beta_i \exp(-\beta_i x), \quad (43)$$

where $f_1(x), f_2(x), \dots, f_n(x)$ denote the density functions of the input exponential distributions. Note, that the mixing procedure can be applied to arbitrary distributions. Using the technique of mixing, one can construct a wide class of distributions. The most commonly used in the applications is a mixture of two exponentials, see Chapter 15 in Cizek, Härdle, and Weron (2005). In the right panel of Figure 5 a pdf of a mixture of two exponentials is plotted together with the pdfs of the mixing laws.

The Laplace transform of (43) is

$$L(t) = \sum_{i=1}^n a_i \frac{\beta_i}{\beta_i + t}, \quad t > -\min_{i=1 \dots n} \{\beta_i\}, \quad (44)$$

yielding the general formula for the k -th raw moment

$$m_k = \sum_{i=1}^n a_i \frac{k!}{\beta_i^k}. \quad (45)$$

The mean is thus $\sum_{i=1}^n a_i \beta_i^{-1}$. The maximum likelihood and method of moments estimators for the mixture of n ($n \geq 2$) exponential distributions can only be evaluated numerically.

Simulation of variates defined by (42) can be performed using the composition approach (Ross, 2002). First generate a random variable I , equal to i with probability a_i , $i = 1, \dots, n$. Then simulate an exponential variate with intensity β_I . Note, that the method is general in the sense that it can be used for any set of distributions F_i 's.

4. Statistical validation techniques

Having a large collection of distributions to choose from we need to narrow our selection to a single model and a unique parameter estimate. The type of the objective loss distribution can be easily selected by comparing the shapes of the empirical and theoretical mean excess functions. The mean excess function, presented in Section 4.1, is based on the idea of conditioning a random variable given that it exceeds a certain level.

Once the distribution class is selected and the parameters are estimated using one of the available methods the goodness-of-fit has to be tested. Probably the most natural approach consists of measuring the distance between the empirical and the fitted analytical distribution function. A group of statistics and tests based on this idea is discussed

in Section 4.2. However, when using these tests we face the problem of comparing a discontinuous step function with a continuous non-decreasing curve. The two functions will always differ from each other in the vicinity of a step by at least half the size of the step. This problem can be overcome by integrating both distributions once, which leads to the so-called limited expected value function introduced in Section 4.3.

4.1. Mean excess function

For a claim amount random variable X , the mean excess function or mean residual life function is the expected payment per claim on a policy with a fixed amount deductible of x , where claims with amounts less than or equal to x are completely ignored:

$$e(x) = E(X - x | X > x) = \frac{\int_x^\infty \{1 - F(u)\} du}{1 - F(x)}. \quad (46)$$

In practice, the mean excess function e is estimated by \hat{e}_n based on a representative sample x_1, \dots, x_n :

$$\hat{e}_n(x) = \frac{\sum_{x_i > x} x_i}{\#\{i : x_i > x\}} - x. \quad (47)$$

Note, that in a financial risk management context, switching from the right tail to the left tail, $e(x)$ is referred to as the expected shortfall (Weron, 2004).

When considering the shapes of mean excess functions, the exponential distribution plays a central role. It has the memoryless property, meaning that whether the information $X > x$ is given or not, the expected value of $X - x$ is the same as if one started at $x = 0$ and calculated $E(X)$. The mean excess function for the exponential distribution is therefore constant. One in fact easily calculates that for this case $e(x) = 1/\beta$ for all $x > 0$.

If the distribution of X is heavier-tailed than the exponential distribution we find that the mean excess function ultimately increases, when it is lighter-tailed $e(x)$ ultimately decreases. Hence, the shape of $e(x)$ provides important information on the sub-exponential or super-exponential nature of the tail of the distribution at hand.

Mean excess functions and first order approximations to the tail for the distributions discussed in Section 3 are given by the following formulas:

- log-normal distribution:

$$e(x) = \frac{\exp\left(\mu + \frac{\sigma^2}{2}\right) \left\{1 - \Phi\left(\frac{\ln x - \mu - \sigma^2}{\sigma}\right)\right\}}{\left\{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)\right\}} - x = \frac{\sigma^2 x}{\ln x - \mu} \{1 + o(1)\},$$

where $o(1)$ stands for a term which tends to zero as $u \rightarrow \infty$;

- exponential distribution:

$$e(x) = \frac{1}{\beta};$$

- Pareto distribution:

$$e(x) = \frac{\lambda + x}{\alpha - 1}, \quad \alpha > 1;$$

- Burr distribution:

$$\begin{aligned} e(x) &= \frac{\lambda^{1/\tau} \Gamma\left(\alpha - \frac{1}{\tau}\right) \Gamma\left(1 + \frac{1}{\tau}\right)}{\Gamma(\alpha)} \cdot \left(\frac{\lambda}{\lambda + x^\tau}\right)^{-\alpha} \cdot \left\{1 - B\left(1 + \frac{1}{\tau}, \alpha - \frac{1}{\tau}, \frac{x^\tau}{\lambda + x^\tau}\right)\right\} - x = \\ &= \frac{x}{\alpha\tau - 1} \{1 + o(1)\}, \quad \alpha\tau > 1, \end{aligned}$$

where $\Gamma(\cdot)$ is the standard gamma function (22) and

$$B(a, b, x) \stackrel{\text{def}}{=} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x y^{a-1} (1-y)^{b-1} dy, \quad (48)$$

is the beta function;

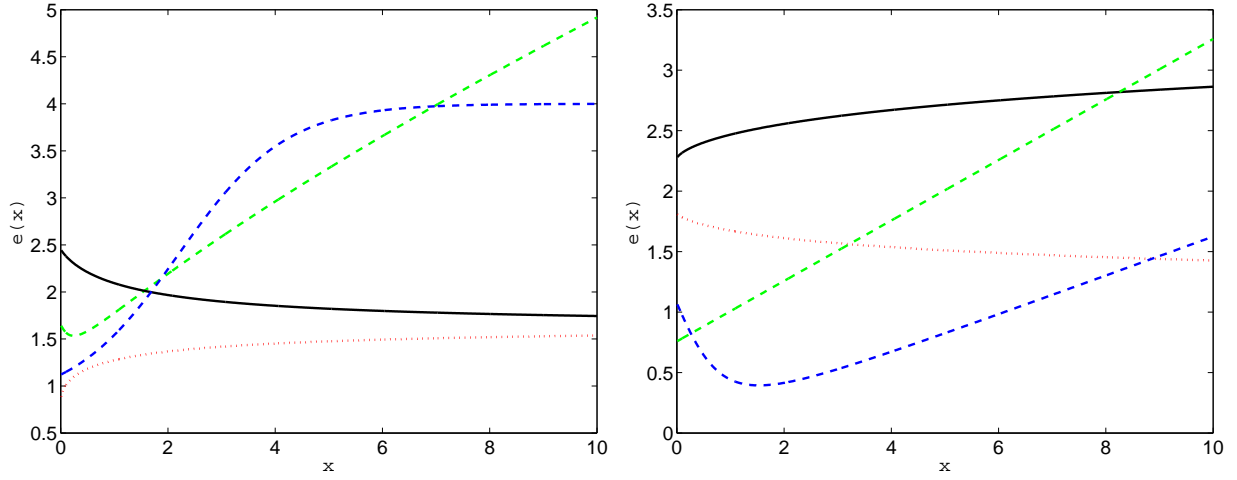


Figure 6: *Left panel:* Shapes of the mean excess function $e(x)$ for the log-normal (green dashed line), gamma with $\alpha < 1$ (red dotted line), gamma with $\alpha > 1$ (black solid line) and a mixture of two exponential distributions (blue long-dashed line). *Right panel:* Shapes of the mean excess function $e(x)$ for the Pareto (green dashed line), Burr (blue long-dashed line), Weibull with $\alpha < 1$ (black solid line) and Weibull with $\alpha > 1$ (red dotted line) distributions.

- Weibull distribution:

$$e(x) = \frac{\Gamma(1 + 1/\tau)}{\beta^{1/\tau}} \left\{ 1 - \Gamma\left(1 + \frac{1}{\tau}, \beta x^\tau\right) \right\} \exp(\beta x^\tau) - x = \frac{x^{1-\tau}}{\beta\tau} \{1 + o(1)\},$$

where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function (35);

- gamma distribution:

$$e(x) = \frac{\alpha}{\beta} \cdot \frac{1 - F(x, \alpha + 1, \beta)}{1 - F(x, \alpha, \beta)} - x = \beta^{-1} \{1 + o(1)\},$$

where $F(x, \alpha, \beta)$ is the gamma distribution function (34);

- mixture of two exponential distributions:

$$e(x) = \frac{\frac{a}{\beta_1} \exp(-\beta_1 x) + \frac{1-a}{\beta_2} \exp(-\beta_2 x)}{a \exp(-\beta_1 x) + (1-a) \exp(-\beta_2 x)}.$$

Selected shapes are also sketched in Figure 6.

4.2. Tests based on the empirical distribution function

A statistics measuring the difference between the empirical $F_n(x)$ and the fitted $F(x)$ distribution function, called an edf statistic, is based on the vertical difference between the distributions. This distance is usually measured either by a supremum or a quadratic norm (D'Agostino and Stephens, 1986).

The most well-known supremum statistic:

$$D = \sup_x |F_n(x) - F(x)|, \quad (49)$$

is known as the Kolmogorov or Kolmogorov-Smirnov statistic. It can also be written in terms of two supremum statistics:

$$D^+ = \sup_x \{F_n(x) - F(x)\} \quad \text{and} \quad D^- = \sup_x \{F(x) - F_n(x)\},$$

where the former is the largest vertical difference when $F_n(x)$ is larger than $F(x)$ and the latter is the largest vertical difference when it is smaller. The Kolmogorov statistic is then given by $D = \max(D^+, D^-)$. A closely related statistic proposed by Kuiper is simply a sum of the two differences, i.e. $V = D^+ + D^-$.

The second class of measures of discrepancy is given by the Cramer-von Mises family

$$Q = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \psi(x) dF(x), \quad (50)$$

where $\psi(x)$ is a suitable function which gives weights to the squared difference $\{F_n(x) - F(x)\}^2$. When $\psi(x) = 1$ we obtain the W^2 statistic of Cramer-von Mises. When $\psi(x) = [F(x)\{1 - F(x)\}]^{-1}$ formula (50) yields the A^2 statistic of Anderson and Darling. From the definitions of the statistics given above, suitable computing formulas must be found. This can be done by utilizing the transformation $Z = F(X)$. When $F(x)$ is the true distribution function of X , the random variable Z is uniformly distributed on the unit interval.

Suppose that a sample x_1, \dots, x_n gives values $z_i = F(x_i)$, $i = 1, \dots, n$. It can be easily shown that, for values z and x related by $z = F(x)$, the corresponding vertical differences in the edf diagrams for X and for Z are equal. Consequently, edf statistics calculated from the empirical distribution function of the z_i 's compared with the uniform distribution will take the same values as if they were calculated from the empirical distribution function of the x_i 's, compared with $F(x)$. This leads to the following formulas given in terms of the order statistics $z_{(1)} < z_{(2)} < \dots < z_{(n)}$:

$$D^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - z_{(i)} \right\}, \quad (51)$$

$$D^- = \max_{1 \leq i \leq n} \left\{ z_{(i)} - \frac{(i-1)}{n} \right\}, \quad (52)$$

$$D = \max(D^+, D^-), \quad (53)$$

$$V = D^+ + D^-, \quad (54)$$

$$W^2 = \sum_{i=1}^n \left\{ z_{(i)} - \frac{(2i-1)}{2n} \right\}^2 + \frac{1}{12n}, \quad (55)$$

$$A^2 = -n - \frac{1}{n}(2i-1) \sum_{i=1}^n \{\log z_{(i)} + \log(1 - z_{(n+1-i)})\} = \quad (56)$$

$$= -n - \frac{1}{n} \sum_{i=1}^n \{(2i-1) \log z_{(i)} + (2n+1-2i) \log(1 - z_{(i)})\}. \quad (57)$$

The general test of fit is structured as follows. The null hypothesis is that a specific distribution is acceptable, whereas the alternative is that it is not:

$$H_0 : F_n(x) = F(x; \theta), \quad H_1 : F_n(x) \neq F(x; \theta),$$

where θ is a vector of known parameters. Small values of the test statistic T are evidence in favor of the null hypothesis, large ones indicate its falsity. To see how unlikely such a large outcome would be if the null hypothesis was true, we calculate the p -value by:

$$p\text{-value} = P(T \geq t), \quad (58)$$

where t is the test value for a given sample. It is typical to reject the null hypothesis when a small p -value is obtained.

However, we are in a situation where we want to test the hypothesis that the sample has a common distribution function $F(x; \theta)$ with unknown θ . To employ any of the edf tests we first need to estimate the parameters. It is important to recognize, however, that when the parameters are estimated from the data, the critical values for the

tests of the uniform distribution (or equivalently of a fully specified distribution) must be reduced. In other words, if the value of the test statistics T is d , then the p -value is overestimated by $P_U(T \geq d)$. Here P_U indicates that the probability is computed under the assumption of a uniformly distributed sample. Hence, if $P_U(T \geq d)$ is small, then the p -value will be even smaller and the hypothesis will be rejected. However, if it is large then we have to obtain a more accurate estimate of the p -value.

Ross (2002) advocates the use of Monte Carlo simulations in this context. First the parameter vector is estimated for a given sample of size n , yielding $\hat{\theta}$, and the edf test statistics is calculated assuming that the sample is distributed according to $F(x; \hat{\theta})$, returning a value of d . Next, a sample of size n of $F(x; \hat{\theta})$ -distributed variates is generated. The parameter vector is estimated for this simulated sample, yielding $\hat{\theta}_1$, and the edf test statistics is calculated assuming that the sample is distributed according to $F(x; \hat{\theta}_1)$. The simulation is repeated as many times as required to achieve a certain level of accuracy. The estimate of the p -value is obtained as the proportion of times that the test quantity is at least as large as d .

An alternative solution to the problem of unknown parameters was proposed by Stephens (1978). The half-sample approach consists of using only half the data to estimate the parameters, but then using the entire data set to conduct the test. In this case, the critical values for the uniform distribution can be applied, at least asymptotically. The quadratic edf tests seem to converge fairly rapidly to their asymptotic distributions (D'Agostino and Stephens, 1986). Although, the method is much faster than the Monte Carlo approach it is not invariant – depending on the choice of the half-samples different test values will be obtained and there is no way of increasing the accuracy.

As a side product, the edf tests supply us with a natural technique of estimating the parameter vector θ . We can simply find such $\hat{\theta}^*$ that minimizes a selected edf statistic. Out of the four presented statistics A^2 is the most powerful when the fitted distribution departs from the true distribution in the tails (D'Agostino and Stephens, 1986). Since the fit in the tails is of crucial importance in most actuarial applications A^2 is the recommended statistic for the estimation scheme.

4.3. Limited expected value function

The limited expected value function L of a claim size variable X , or of the corresponding cdf $F(x)$, is defined by

$$L(x) = E\{\min(X, x)\} = \int_0^x y dF(y) + x\{1 - F(x)\}, \quad x > 0. \quad (59)$$

The value of the function L at point x is equal to the expectation of the cdf $F(x)$ truncated at this point. In other words, it represents the expected amount per claim retained by the insured on a policy with a fixed amount deductible of x . The empirical estimate is defined as follows:

$$\hat{L}_n(x) = \frac{1}{n} \left(\sum_{x_j < x} x_j + \sum_{x_j \geq x} x \right). \quad (60)$$

In order to fit the limited expected value function L of an analytical distribution to the observed data, the estimate \hat{L}_n is first constructed. Thereafter one tries to find a suitable analytical cdf F , such that the corresponding limited expected value function L is as close to the observed \hat{L}_n as possible.

The limited expected value function has the following important properties:

1. the graph of L is concave, continuous and increasing;
2. $L(x) \rightarrow E(X)$, as $x \rightarrow \infty$;
3. $F(x) = 1 - L'(x)$, where $L'(x)$ is the derivative of the function L at point x ; if F is discontinuous at x , then the equality holds true for the right-hand derivative $L'(x+)$.

A reason why the limited expected value function is a particularly suitable tool for our purposes is that it represents the claim size distribution in the monetary dimension. For example, we have $L(\infty) = E(X)$ if it exists. The cdf F , on the other hand, operates on the probability scale, i.e. takes values between 0 and 1. Therefore, it is usually difficult to see, by looking only at $F(x)$, how sensitive the price for the insurance – the premium – is to changes in the values of F , while the limited expected value function shows immediately how different parts of the claim size cdf contribute to the premium (see Chapter 19 in Cížek, Härdle, and Weron (2005) for information on various premium calculation

principles). Apart from curve-fitting purposes, the function L will turn out to be a very useful concept in dealing with deductibles in Chapter 19 in Čižek, Härdle, and Weron (2005). It is also worth mentioning, that there exists a connection between the limited expected value function and the mean excess function:

$$E(X) = L(u) + P(X > u)e(u). \quad (61)$$

The limited expected value functions for all distributions considered in this chapter are given by:

- log-normal distribution:

$$L(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \Phi\left(\frac{\ln x - \mu - \sigma^2}{\sigma}\right) + x \left\{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)\right\};$$

- exponential distribution:

$$L(x) = \frac{1}{\beta} \{1 - \exp(-\beta x)\};$$

- Pareto distribution:

$$L(x) = \frac{\lambda - \lambda^\alpha (\lambda + x)^{1-\alpha}}{\alpha - 1};$$

- Burr distribution:

$$L(x) = \frac{\lambda^{1/\tau} \Gamma\left(\alpha - \frac{1}{\tau}\right) \Gamma\left(1 + \frac{1}{\tau}\right)}{\Gamma(\alpha)} B\left(1 + \frac{1}{\tau}, \alpha - \frac{1}{\tau}; \frac{x^\tau}{\lambda + x^\tau}\right) + x \left(\frac{\lambda}{\lambda + x^\tau}\right)^\alpha;$$

- Weibull distribution:

$$L(x) = \frac{\Gamma(1 + 1/\tau)}{\beta^{1/\tau}} \Gamma\left(1 + \frac{1}{\tau}, \beta x^\tau\right) + x e^{-\beta x^\tau};$$

- gamma distribution:

$$L(x) = \frac{\alpha}{\beta} F(x, \alpha + 1, \beta) + x \{1 - F(x, \alpha, \beta)\};$$

- mixture of two exponential distributions:

$$L(x) = \frac{a}{\beta_1} \{1 - \exp(-\beta_1 x)\} + \frac{1-a}{\beta_2} \{1 - \exp(-\beta_2 x)\}.$$

From the curve-fitting point of view the use of the limited expected value function has the advantage, compared with the use of the cdfs, that both the analytical and the corresponding observed function \hat{L}_n , based on the observed discrete cdf, are continuous and concave, whereas the observed claim size cdf F_n is a discontinuous step function. Property (3) implies that the limited expected value function determines the corresponding cdf uniquely. When the limited expected value functions of two distributions are close to each other, not only are the mean values of the distributions close to each other, but the whole distributions as well.

5. Applications

In this section we illustrate some of the methods described earlier in the chapter. We conduct the analysis for two datasets. The first is the PCS (Property Claim Services, see Insurance Services Office Inc. (ISO) web site: www.iso.com/products/2800/prod2801.html) dataset covering losses resulting from natural catastrophic events in USA exceeding 5 million dollars that occurred between 1990 and 1999. The second is the Danish fire losses dataset, which concerns major fire losses in Danish Krone (DKK) that occurred between 1980 and 1990 and were

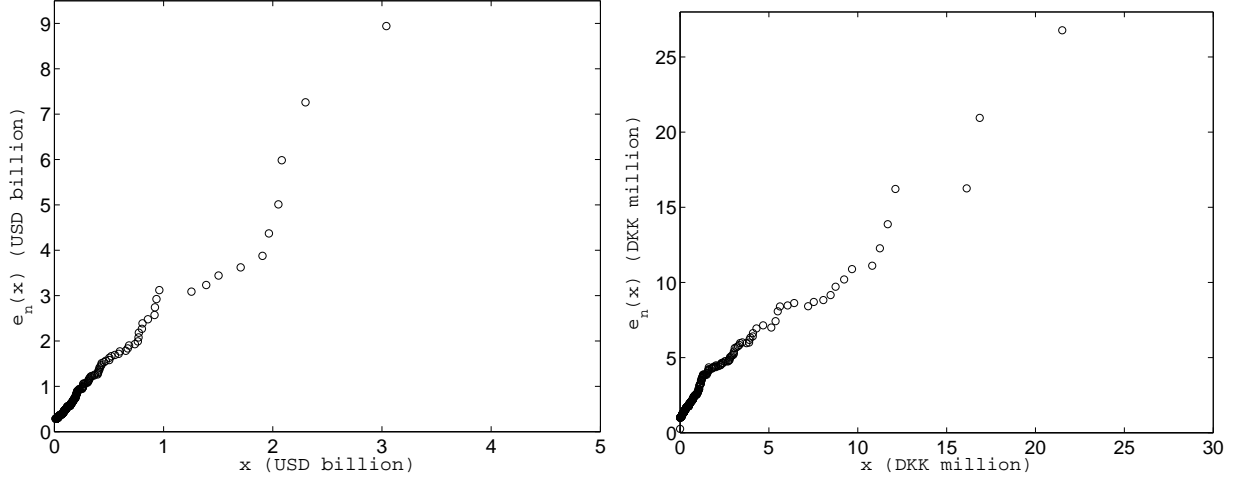


Figure 7: The empirical mean excess function $\hat{e}_n(x)$ for the PCS catastrophe data (*left panel*) and the Danish fire data (*right panel*).

recorded by Copenhagen Re. Here we consider only losses in profits. The overall fire losses were analyzed by Embrechts, Klüppelberg, and Mikosch (1997).

The Danish fire losses dataset has been already adjusted for inflation. However, the PCS dataset consists of raw data. Since the data have been collected over a considerable period of time, it is important to bring the values onto a common basis by means of a suitably chosen index. The choice of the index depends on the line of insurance. For example, an index of the cost of construction prices may be suitable for fire and other property insurance, an earnings index for life and accident insurance, and a general price index may be appropriate when a single index is required for several lines or for the whole portfolio. Here we adjust the PCS dataset using the Consumer Price Index provided by the U.S. Department of Labor. Note, that the same raw catastrophe data, however, adjusted using the discount window borrowing rate that refers to the simple interest rate at which depository institutions borrow from the Federal Reserve Bank of New York was analyzed by Burnecki, Härdle, and Weron (2004). A related dataset containing the national and regional PCS indices for losses resulting from catastrophic events in USA was studied by Burnecki, Kukla, and Weron (2000).

As suggested in the proceeding section we first look for the appropriate shape of the distribution. To this end we plot the empirical mean excess functions for the analyzed data sets, see Figure 7. Both in the case of PCS natural catastrophe losses and Danish fire losses the data show a super-exponential pattern suggesting a log-normal, Pareto or Burr distribution as most adequate for modeling. Hence, in the sequel we calibrate these three distributions.

We apply two estimation schemes: maximum likelihood and A^2 statistics minimization. Out of the three fitted distributions only the log-normal has closed form expressions for the maximum likelihood estimators. Parameter calibration for the remaining distributions and the A^2 minimization scheme is carried out via a simplex numerical optimization routine. A limited simulation study suggests that the A^2 minimization scheme tends to return lower values of all edf test statistics than maximum likelihood estimation. Hence, it is exclusively used for further analysis.

The results of parameter estimation and hypothesis testing for the PCS loss amounts are presented in Table 1. The Burr distribution with parameters $\alpha = 0.4801$, $\lambda = 3.9495 \cdot 10^{16}$, and $\tau = 2.1524$ yields the best results and passes all tests at the 3% level. The log-normal distribution with parameters $\mu = 18.3806$ and $\sigma = 1.1052$ comes in second, however, with an unacceptable fit as tested by the Anderson-Darling statistics. As expected, the remaining distributions presented in Section 3 return even worse fits. Thus, here we suggest to choose the Burr distribution as a model for the PCS loss amounts. In the left panel of Figure 8 we present the empirical and analytical limited expected value functions for the three fitted distributions. The plot justifies the choice of the Burr distribution.

Interestingly, apart from the two very extreme observations (Hurricane Andrew and the Northridge Earthquake), the Pareto law provides a very good fit as well. This can be seen in Figure 9 where the Pareto probability plot of the PCS catastrophe data is displayed. Recall that the probability plot is constructed by first ordering the observations

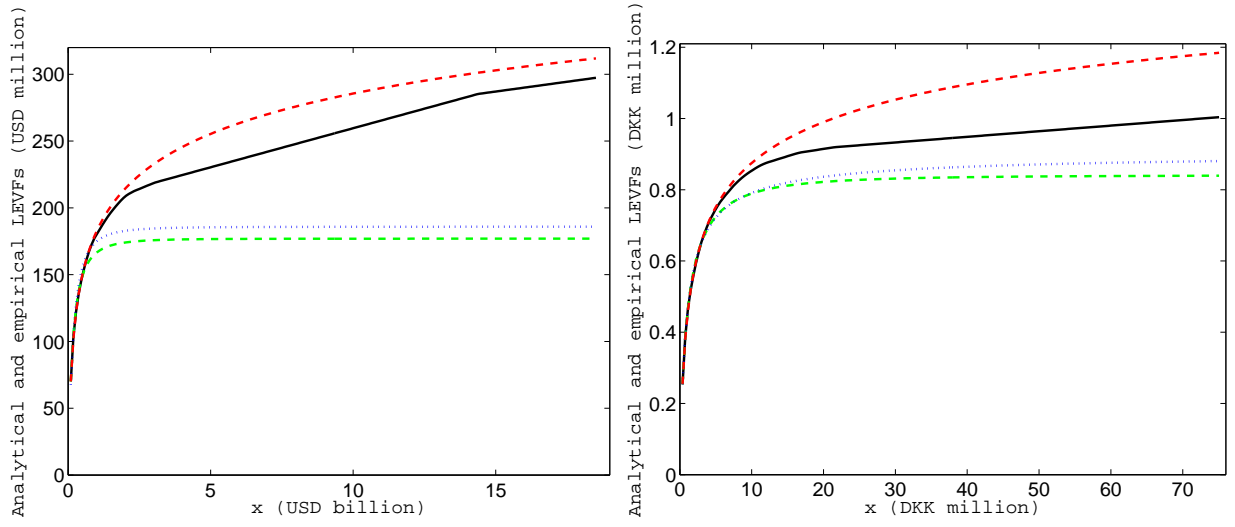


Figure 8: The empirical (black solid line) and analytical limited expected value functions (LEVFs) for the log-normal (green dashed line), Pareto (blue dotted line), and Burr (red long-dashed line) distributions for the PCS catastrophe data (*left panel*) and the Danish fire data (*right panel*).

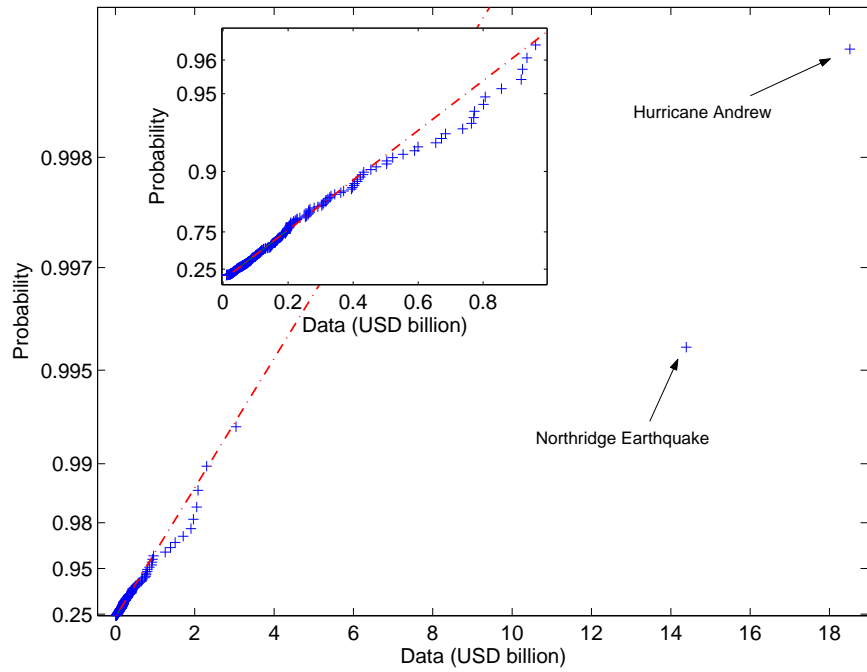


Figure 9: Pareto probability plot of the PCS catastrophe data. Apart from the two very extreme observations (Hurricane Andrew and Northridge Earthquake) the points (pluses) more or less constitute a straight line, validating the choice of the Pareto distribution. The inset is a magnification of the bottom left part of the original plot.

Table 1: Parameter estimates obtained via the A^2 minimization scheme and test statistics for the catastrophe loss amounts. The corresponding p -values based on 1000 simulated samples are given in parentheses.

Distributions:		log-normal	Pareto	Burr
Parameters:		$\mu=18.3806$ $\sigma=1.1052$	$\alpha=3.4081$ $\lambda=4.4767 \cdot 10^8$	$\alpha=0.4801$ $\lambda=3.9495 \cdot 10^{16}$ $\tau=2.1524$
Tests:	D	0.0440 (0.032)	0.1049 (<0.005)	0.0366 (0.072)
	V	0.0786 (0.028)	0.1692 (<0.005)	0.0703 (0.031)
	W^2	0.1353 (0.012)	0.7042 (<0.005)	0.0626 (0.068)
	A^2	1.8606 (<0.005)	6.1160 (<0.005)	0.5097 (0.034)

Table 2: Parameter estimates obtained via the A^2 minimization scheme and test statistics for the fire loss amounts. The corresponding p -values based on 1000 simulated samples are given in parentheses.

Distributions:		log-normal	Pareto	Burr
Parameters:		$\mu=12.6645$ $\sigma=1.3981$	$\alpha=1.7439$ $\lambda=6.7522 \cdot 10^5$	$\alpha=0.8804$ $\lambda=8.4202 \cdot 10^6$ $\tau=1.2749$
Tests:	D	0.0381 (<0.005)	0.0471 (<0.005)	0.0387 (<0.005)
	V	0.0676 (<0.005)	0.0779 (<0.005)	0.0724 (<0.005)
	W^2	0.0921 (0.055)	0.2119 (<0.005)	0.1117 (<0.005)
	A^2	0.7567 (0.022)	1.9097 (<0.005)	0.6999 (<0.005)

x_1, \dots, x_n from the smallest to the largest: $x_{(1)} \leq \dots \leq x_{(n)}$ and next plotting them against their observed cumulative frequency, i.e. the points (pluses in Fig. 9) correspond to the pairs $(x_{(i)}, F^{-1}([i - 0.5]/n))$, for $i = 1, \dots, n$. If the hypothesized distribution F (here Pareto) adequately describes the data, the plotted points fall approximately along a straight line (Burnecki and Weron, 2008). The procedure of trimming the top 1-5% of the data before calibration is known as robust estimation. Chernobai et al. (2006) showed that trimming leads to accepting the Pareto law as a model of PCS loss sizes.

The results of parameter estimation and hypothesis testing for the Danish fire loss amounts are presented in Table 2. The log-normal distribution with parameters $\mu = 12.6645$ and $\sigma = 1.3981$ returns the best results. It is the only distribution that passes any of the four applied tests (D , V , W^2 , and A^2) at a reasonable level. The Burr and Pareto laws yield worse fits as the tails of the edf are lighter than power-law tails. As expected, the remaining distributions presented in Section 3 return even worse fits. In the right panel of Figure 8 we depict the empirical and analytical limited expected value functions for the three fitted distributions. Unfortunately, no definitive conclusions can be drawn regarding the choice of the distribution. Hence, we suggest to use the log-normal distribution as a model for the Danish fire loss amounts.

References

- Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A Guide to Simulation*, Springer-Verlag, New York.
- Burnecki, K., Härdle, W., and Weron, R. (2004). Simulation of risk processes, in J. Teugels, B. Sundt (eds.) *Encyclopedia of Actuarial Science*, Wiley, Chichester.
- Burnecki, K., Kukla, G., and Weron, R. (2000). Property insurance loss distributions, *Physica A* **287**: 269-278.
- Burnecki, K. and Weron, R. (2008). Visualization Tools for Insurance Risk Processes, in Ch. Chen, W. Härdle, A. Unwin (eds.) *Handbook of Data Visualization*, Springer, Berlin, 899-920.
- Chernobai, A., Burnecki, K., Rachev, S.T., Trück, S., and Weron, R. (2006). Modelling catastrophe claims with left-truncated severity distributions, *Computational Statistics* **21**: 537-555.
- Cižek, P., Härdle, W., and Weron, R., eds. (2005). *Statistical Tools for Finance and Insurance*, Springer-Verlag, Berlin.
- D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, New York.
- Daykin, C.D., Pentikainen, T., and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*, Chapman, London.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer.
- Hogg, R. and Klugman, S. A. (1984). *Loss Distributions*, Wiley, New York.
- Jöhnk, M. D. (1964). Erzeugung von Betaverteilten und Gammaverteilten Zufallszahlen, *Metrika* **8**: 5-15.
- Klugman, S. A., Panjer, H.H., and Willmot, G.E. (1998). *Loss Models: From Data to Decisions*, Wiley, New York.
- L'Ecuyer, P. (2004). Random Number Generation, in J. E. Gentle, W. Härdle, Y. Mori (eds.) *Handbook of Computational Statistics*, Springer, Berlin, 35-70.
- Panjer, H.H. and Willmot, G.E. (1992). *Insurance Risk Models*, Society of Actuaries, Chicago.
- Ross, S. (2002). *Simulation*, Academic Press, San Diego.
- Stephens, M. A. (1978). On the half-sample method for goodness-of-fit, *Journal of the Royal Statistical Society B* **40**: 64-70.
- Weron, R. (2004). Computationally Intensive Value at Risk Calculations, in J. E. Gentle, W. Härdle, Y. Mori (eds.) *Handbook of Computational Statistics*, Springer, Berlin, 911-950.